

Blaise Aguera y Arcas

Cómo las computadoras aprenden a ser creativas

Coordino un equipo de Google que trabaja en inteligencia artificial; es decir, en la disciplina de ingeniería para fabricar computadoras y dispositivos capaces de hacer algunas cosas que hacen los cerebros. Y esto hace que nos interese mucho el cerebro real, la neurociencia y, especialmente, nos interesan las cosas que hace nuestro cerebro que todavía son muy superiores al rendimiento de las computadoras.

Históricamente, una de esas áreas ha sido la percepción, el proceso por el cual las cosas que hay en el mundo, sonidos e imágenes, pueden convertirse en conceptos en la mente. Esto es esencial para nuestro propio cerebro, y también es muy útil en una máquina. Nuestro equipo hace algoritmos de percepción computacional que permiten encontrar imágenes en Google en función de lo que hay en ellas. La otra cara de la percepción es la creatividad: convertir un concepto en algo que hay en el mundo. Así que en el último año, nuestro trabajo en percepción computacional ha conectado de forma inesperada con el mundo de la creatividad y el arte computacionales.

Creo que Miguel Ángel tenía una visión aguda en esta doble relación entre la percepción y la creatividad. Esta es una de sus célebres citas: "Cada bloque de piedra tiene una estatua en su interior, y el trabajo del escultor es descubrirla". Pienso que Miguel Ángel quería decir que creamos al percibir, y que la propia percepción es un acto de imaginación y es la materia de la creatividad.

El órgano que crea todo el pensamiento, la percepción y la imaginación, por supuesto, es el cerebro. Y me gustaría empezar con un poquito de historia sobre lo que sabemos del cerebro. Porque a diferencia de, por ejemplo, el corazón o los intestinos, realmente no se puede decir mucho de un cerebro con solo mirarlo, al menos a simple vista. Los primeros anatomistas que exploraron cerebros dieron a las estructuras superficiales de esta cosa todo tipo de nombres de fantasía, como hipocampo, que significa "pequeño camarón". Pero, por supuesto, eso no nos dice mucho de lo que realmente sucede en el interior.

La primera persona que, creo, desarrolló una visión de lo que ocurría en el cerebro fue el gran neuroanatomista español, Santiago Ramón y Cajal, en el siglo XIX, que usó la microscopía y tinciones especiales para poder marcar selectivamente en un contraste muy alto las células individuales en el cerebro, para empezar a entender sus morfologías. Y estos son los tipos de dibujos que hizo de las neuronas en el siglo XIX.

Esto es de un cerebro de un pájaro. Y ya ven esta increíble variedad de diferentes tipos de células, incluso la propia teoría celular era bastante nueva en este momento. Y estas estructuras, estas células que tienen estas arborizaciones, estas ramas que pueden alcanzar muy largas distancias, esto era muy novedoso en esa época. Nos recuerdan, por supuesto, a los cables. Eso podría haber sido obvio para algunos en el siglo XIX; las revoluciones

del cableado y de la electricidad se estaban iniciando. Pero en muchos sentidos, los dibujos de microanatomía de Ramón y Cajal, como éste, todavía son, en cierto modo, insuperables.

Un siglo más tarde todavía tratamos de terminar el trabajo que empezó Ramón y Cajal. Estos son los datos brutos de nuestros colaboradores del Instituto Max Planck de Neurociencia. Nuestros colaboradores han tomado imágenes de trozos de tejido cerebral. La muestra completa aquí es de 1 mm cúbico de tamaño, y les estoy mostrando un trozo muy pequeño. Esa barra de la izquierda es de aproximadamente una micra. Las estructuras visibles son las mitocondrias del tamaño de las bacterias. Y estos son cortes consecutivos de este diminuto bloque de tejido. solo a efectos de comparación, el diámetro de una hebra promedio de pelo es de unas 100 micras. Así que vemos algo mucho, mucho más pequeño que una sola hebra de cabello.

Y a partir de este tipo de rebanadas de microscopía electrónica de serie, se pueden hacer reconstrucciones en 3D de las neuronas con este aspecto. Son casi del mismo estilo que las de Ramón y Cajal. Solo se iluminaron unas pocas neuronas, porque, de lo contrario, no se podría ver nada aquí. Estaría tan lleno, tan pleno de estructuras, de cableado de todas las neuronas conectadas una a otra.

Así Ramón y Cajal se adelantó un poco a su tiempo, y al progreso en la comprensión del cerebro avanzando lentamente a lo largo de las siguientes décadas. Pero sabíamos que las neuronas usan electricidad y por la Segunda Guerra Mundial, la tecnología avanzó lo suficiente como para empezar a hacer experimentos eléctricos reales con neuronas vivas para comprender mejor cómo funcionaban. En ese mismo momento se desarrollaban las computadoras con la idea de modelar el cerebro, de "máquinas inteligentes", como decía Alan Turing, uno de los padres de la informática.

Warren McCulloch y Walter Pitts miraron el dibujo de Ramón y Cajal de la corteza visual, que muestro aquí. Esta es la corteza que procesa las imágenes que provienen del ojo. Y para ellos, esto parecía un diagrama de circuito. Y hay gran cantidad de detalles en el diagrama de circuito de McCulloch y Pitts que no están del todo bien. Pero esta idea básica de que la corteza visual funciona como una serie de elementos computacionales que pasan una información de uno al siguiente en cascada, es esencialmente correcta.

Hablemos por un momento de lo que tendría que hacer un modelo para procesar la información visual. La tarea básica de la percepción es tomar una imagen como esta y decir: "Eso es un pájaro" que es algo muy simple de ver con nuestro cerebro. Pero todos Uds. deben entender que para una computadora esto era prácticamente imposible hace pocos años. Con el paradigma de la computación clásica esta tarea no es fácil de hacer.

Entonces lo que pasa entre los píxeles, entre la imagen del ave y la palabra "pájaro" es esencialmente un conjunto de neuronas conectadas entre sí en una red neuronal, como la que diagramo aquí. Esta red neuronal podría ser biológica, en nuestras cortezas visuales,

o, en la actualidad, podemos modelar este tipo de redes neuronales en la computadora. Y mostraré qué aspecto tienen.

Así que los píxeles se puede pensar como una primera capa de neuronas, y así es, de hecho, como funciona el ojo, eso son las neuronas de la retina. Y después avanzan de una capa a la otra, y luego a otra capa de neuronas, todas conectadas por sinapsis de diferentes pesos. El comportamiento de esta red se caracteriza por las fortalezas de todas esas sinapsis. Estas caracterizan las propiedades computacionales de esta red. Y al final una neurona o un pequeño grupo de neuronas da la luz, diciendo, "pájaro".

Ahora voy a representar esas tres cosas: los píxeles de entrada, las sinapsis en la red neuronal, y el pájaro, la salida, con tres variables: X , W e Y . Hay tal vez un millón o más de X , un millón de píxeles en la imagen. Hay miles de millones o billones de W , que representan los pesos de todas estas sinapsis en la red neuronal. Y hay un número muy pequeño de Y , de salidas que tiene esa red. "Pájaro" son solo seis letras, ¿verdad? Así que vamos a suponer que esto es solo una fórmula simple, $X \times W = Y$. Pongo la multiplicación entre comillas porque lo que realmente pasa allí, por supuesto, es una serie muy complicada de operaciones matemáticas.

Esa es una ecuación. Hay tres variables. Y todos sabemos que si uno tiene una ecuación, puede resolver una variable conociendo las otras dos. Así que el problema de la inferencia, es decir, averiguar que la imagen de un pájaro es un pájaro, es este: Y es la desconocida y W y X las conocidas. Se conoce la red neuronal, y también los píxeles. Como se puede ver, en realidad, es un problema relativamente sencillo. Se multiplica dos veces tres y ya está. Les voy a mostrar una red neuronal artificial que hemos construido recientemente, haciendo exactamente eso.

Esto se ejecuta en tiempo real en un teléfono móvil, y eso es, por supuesto, sorprendente en sí mismo, que los teléfonos móviles puedan hacer tantos millones y billones de operaciones por segundo. Lo que ven es un teléfono que analiza imágenes de un pájaro una tras otra. Y, de hecho, no solo dice: "Sí, es un pájaro" sino que identifica las especies de pájaros con una red de este tipo. Así que en ese cuadro, la X y la W son conocidas, y la Y es la desconocida. Estoy pasando por alto la parte más difícil, por supuesto, que es cómo demonios podemos averiguar la W , cómo puede el cerebro hacerlo. ¿Cómo podríamos llegar a aprender un modelo de este tipo?

Este proceso de aprendizaje de despejar W , si lo hacemos con una simple ecuación en la que pensamos en ellos como números, sabemos exactamente cómo hacer eso: $6 = 2 \times W$, así, se divide por dos y ya está. El problema estriba en este operador. Por lo tanto, la división, hemos usado la división por ser la inversa de la multiplicación, pero, como acabo de decir, la multiplicación tiene algo de mentira aquí. Esta es una operación muy complicada, nada lineal; que no tiene inversa. Así que tenemos que encontrar una manera de resolver la ecuación sin un operador de división. Y la manera de hacerlo es bastante sencilla. Vamos a aplicar un pequeño truco de álgebra, y a mover el 6 hacia el lado derecho

de la ecuación. Ahora, todavía usamos la multiplicación. Y el cero... pensémoslo como un error. Es decir, si hemos resuelto la W de la manera correcta, luego el error será el cero. Y si no lo hacemos del todo bien, el error será mayor que cero.

Así que ahora solo podemos hacer conjeturas para minimizar el error, y en eso las computadoras son muy buenas. Así que ya hemos hecho una aproximación inicial: ¿Y si $W = 0$? Entonces el error es 6. ¿Qué pasa si $W = 1$? El error es 4. Y luego, la computadora puede correr una especie de Marco Polo, y reducir el error cercano a cero. A medida que lo hace, se logra aproximaciones sucesivas a W . Por lo general, nunca llega allí, pero tras una docena de pasos, estamos en $W = 2.999$, que es lo suficientemente aproximado. Y este es el proceso de aprendizaje.

Así que recuerden que hemos tomado muchas X e Y conocidas para resolver la W por medio de un proceso iterativo. De la misma manera como lo hacemos en nuestro propio aprendizaje. Tenemos muchas, muchas imágenes de bebés y nos dicen: "Esto es un pájaro, esto no es un pájaro". Y con el tiempo, a través de iteración, resolvemos W , lo resolvemos para esas conexiones neuronales.

Así que ahora ya tenemos despejada la X , y la W para resolver Y ; eso todos los días, percepción rápida. Entendemos cómo podemos resolver W , esto es aprendizaje, que es mucho más difícil, porque tenemos que minimizar errores, usando mucho ejemplos para el entrenamiento.

Y hace un año, Alex Mordvintsev, de nuestro equipo, decidió experimentar qué sucede si intentamos resolver X , con una W e Y conocidas. En otras palabras, se sabe que es un pájaro, y se cuenta con una red neuronal entrenada en aves, pero ¿y la imagen de un pájaro? Usando el mismo procedimiento de minimización de errores, uno puede hacer eso con la red entrenada para reconocer aves, y el resultado es... una imagen de aves. Esta es una imagen de aves generada en su totalidad por una red neuronal entrenada para reconocer aves, simplemente resolviendo X , en lugar de resolver Y , haciéndolo de forma iterativa.

He aquí otro ejemplo divertido. Este fue un trabajo realizado por Mike Tyka en nuestro grupo, llamado "animal del desfile". Me recuerda algo a las obras de arte de William Kentridge, donde él hace bocetos, los borra, hace bocetos, los borra y crea una película de esta manera. En este caso, Mike varía Y sobre el espacio de diferentes animales, en una red diseñada para reconocer y distinguir animales diferentes unos de otros. Y se obtiene esta extraña metamorfosis de un animal a la Escher.

Aquí él y Alex han intentado reducir las Y a un espacio de solo dos dimensiones, logrando un mapa fuera del espacio de todas las cosas reconocido por esta red. Realizar este tipo de síntesis o la generación de las imágenes sobre toda la superficie, variando Y sobre la superficie, se hace una especie de mapa, un mapa visual de todas las cosas que la red sabe reconocer. Los animales están todos aquí; el armadillo está justo aquí.

Se puede hacer esto con otros tipos de redes. Esta es una red diseñada para reconocer caras, para distinguir una cara de otra. Y aquí ponemos en Y una que dice "yo", mis propios parámetros cara. Y cuando esto resuelve X, genera esta imagen de mí, alocada, tipo cubista, psicodélica, como un cuadro surrealista, desde múltiples puntos de vista a la vez. Aparecen múltiples puntos de vista a la vez porque la red está diseñada para descartar la ambigüedad de una cara en una postura u otra, con un tipo de luz, u otro. Al hacer este tipo de reconstrucción, si no se usa alguna imagen de guía o estadísticas de guía, entonces se obtiene una confusión de diferentes puntos de vista, porque es ambigua. Esto es lo que sucede si Alex usa su propia cara imagen como guía durante ese proceso de optimización para reconstruir mi propia cara. Así se puede ver que no es perfecto. Todavía hay mucho trabajo por hacer sobre cómo mejorar el proceso de optimización. Pero ya se empieza a ver algo más parecido a una cara coherente, usando mi propia cara como guía.

No tiene que comenzar con un lienzo en blanco o con ruido blanco. Cuando se está resolviendo X, se puede comenzar con una X, que en sí es ya una imagen. En eso consiste esta pequeña demostración. Esta es una red diseñada para categorizar todo objeto, estructuras hechas por humanos, animales... Aquí empezamos con una imagen de las nubes, y la optimizamos, básicamente, esta red averigua qué se ve en las nubes. Y cuanto más tiempo uno pasa mirando, más cosas también se verán en las nubes. También es posible usar la red para alucinar, obteniendo cosas bastante locas.

O, Mike ha hecho otros experimentos donde se detiene la imagen de la nube, la alucina, la amplía, la alucina, la amplía... Y de esta manera, se obtiene una especie de estado de fuga de la red, supongo, o una especie de asociación libre, en el que la red se come su propia cola. Así que cada imagen es ahora la base para "¿Qué pienso que veré ahora? ¿Qué pienso que veré ahora? ¿Qué pienso que veré ahora?"

Mostré esto por primera vez en público a un grupo en una conferencia en Seattle llamada "Educación Superior", esto fue justo después de que la marihuana fuera legalizada.

Así que me gustaría terminar rápidamente con solo señalar que esta tecnología no está limitada. He mostrado ejemplos puramente visuales porque son muy divertidos. Pero no es una tecnología puramente visual. Nuestro artista colaborador, Ross Goodwin, tiene experimentos que implican a una cámara tomando una foto, y luego una computadora en su mochila, escribe un poema usando redes neuronales, basado en el contenido de la imagen. Y que la red neuronal de poesía ha sido entrenada en un gran corpus de poesía del siglo XX. Y la poesía es, ya saben, creo que no está tan mal, en realidad.

Para concluir, creo que Miguel Ángel tenía razón; la percepción y la creatividad están conectadas muy íntimamente. Acabamos de ver redes neuronales totalmente capacitadas para discriminar, o para reconocer cosas diferentes en el mundo, que pueden ejecutarse a la inversa, para generar nuevas cosas. Algo que me sugiere esto no es solo que Miguel Ángel realmente vio la escultura en los bloques de piedra, sino que cualquier criatura,

cualquier ser, cualquier alienígena que es capaz de percibir actos de ese tipo también es capaz de crear porque en ambos casos se usa la misma maquinaria.

Además, creo que la percepción y la creatividad no son absolutamente únicamente humanas. Ya tenemos modelos computacionales que pueden hacer exactamente estas cosas. Y que no debería ser sorprendente; el cerebro es computacional.

Y, finalmente, la computación empezó como un ejercicio de diseño de máquinas inteligentes. Fue modelada siguiendo la idea de cómo podríamos hacer máquinas inteligentes. Y, finalmente, se están empezando a cumplir algunas de las promesas de aquellos pioneros, como Turing, von Neumann McCulloch y Pitts. Y creo que la informática no es solo contabilidad o jugar al Candy Crush y esas cosas. Desde el principio, se diseñó siguiendo el modelo de nuestra mente. Y eso nos da la capacidad de comprender mejor nuestra propia mente y de ampliarla.

Muchas gracias.